Appendix

A. Proof of Lemma 1

Assumption (Assumption from [17]). *(1) Given an arbitrary ϕ, $q_{tj}(x|\mathbf{z}_t;\phi) > 0$ for all $x \in \mathbb{R}$, $j = 1, \cdots, p$, and $\mathbf{z}_t \in \mathbb{R}^d$, $t = C+1, \cdots, T$. (2) Given an arbitrary θ, $Q(\cdot, \mathbf{z}_t; \theta_{tj})$ is invertible and differentiable for all $j = 1, \cdots, p$ and $\mathbf{z}_t \in \mathbb{R}^d$, $t = C+1, \cdots, T$.*

Lemma. *For an arbitrary ϕ, let $\theta^*(\phi) \in \arg\min_\theta \mathbb{E}_p \mathbb{E}_q \sum_{t=C+1}^{T} \sum_{j=1}^{p} \int_0^1 \rho_\alpha\big(\mathbf{x}_{tj} - Q(\alpha, \mathbf{z}_t; \theta_{tj})\big) d\alpha$, where $\theta^*(\phi) = (\theta^*_{C+1}(\phi), \cdots, \theta^*_T(\phi))$, $\theta^*_t(\phi) = (\theta^*_{t1}(\phi), \cdots, \theta^*_{tp}(\phi))$, $t = C+1, \cdots, T$, and p and q denote $p(\mathbf{x}_{C+1:T}|\mathbf{x}_{1:C})$ and $q(\mathbf{z}_{C+1:T}|\mathbf{x}_{1:T};\phi)$, respectively. Then,*

$$\mathcal{KL}\left(\int \prod_{t=C+1}^{T} \prod_{j=1}^{p} \frac{d}{d\mathbf{x}_{tj}} Q^{-1}(\mathbf{x}_{tj}, \mathbf{z}_t; \theta^*_{tj}(\phi)) q(\mathbf{z}_{C+1:T}|\mathbf{x}_{1:C};\phi) d\mathbf{z}_{C+1:T} \;\middle\|\; p(\mathbf{x}_{C+1:T}|\mathbf{x}_{1:C}) \right) = 0,$$

Proof.

$$\begin{aligned}
\min_\theta \quad & \mathbb{E}_{p(\mathbf{x}_{C+1:T}|\mathbf{x}_{1:C})} \mathbb{E}_{q(\mathbf{z}_{C+1:T}|\mathbf{x}_{1:T};\phi)} \sum_{t=C+1}^{T} \sum_{j=1}^{p} \int_0^1 \rho_\alpha\big(\mathbf{x}_{tj} - Q(\alpha, \mathbf{z}_t; \theta_{tj})\big) d\alpha \\
&= \mathbb{E}_{q(\mathbf{x}_{C+1:T}|\mathbf{z}_{C+1:T};\phi) q(\mathbf{z}_{C+1:T}|\mathbf{x}_{1:C};\phi)} \sum_{t=C+1}^{T} \sum_{j=1}^{p} \int_0^1 \rho_\alpha\big(\mathbf{x}_{tj} - Q(\alpha, \mathbf{z}_t; \theta_{tj})\big) d\alpha \\
&= \mathbb{E}_{q(\mathbf{z}_{C+1:T}|\mathbf{x}_{1:C};\phi)} \sum_{t=C+1}^{T} \sum_{j=1}^{p} \mathbb{E}_{q(\mathbf{x}_{tj}|\mathbf{z}_t;\phi)} \int_0^1 \rho_\alpha\big(\mathbf{x}_{tj} - Q(\alpha, \mathbf{z}_t; \theta_{tj})\big) d\alpha
\end{aligned}$$

Therefore, for all $t = C+1, \cdots, T$, and $j = 1, \cdots, p$,

$$\theta^*_{tj}(\phi) = \arg\min_{\theta_{tj}} \mathbb{E}_{q(\mathbf{x}_{tj}|\mathbf{z}_t;\phi)} \int_0^1 \rho_\alpha\big(\mathbf{x}_{tj} - Q(\alpha, \mathbf{z}_t; \theta_{tj})\big) d\alpha$$

and it is proper relative to $F_{tj}(\mathbf{x}_{tj}|\mathbf{z}_t;\phi)$ (CRPS loss). It implies that $Q^{-1}(\mathbf{x}_{tj}, \mathbf{z}_t; \theta^*_{tj}(\phi)) = F_{tj}(\mathbf{x}_{tj}|\mathbf{z}_t;\phi)$, and $\frac{d}{d\mathbf{x}_{tj}} Q^{-1}(\mathbf{x}_{tj}|\mathbf{z}_t; \theta^*_{tj}(\phi)) = q_{tj}(\mathbf{x}_{tj}|\mathbf{z}_t;\phi)$, by assumption.

It follows that

$$\begin{aligned}
& \int \prod_{t=C+1}^{T} \prod_{j=1}^{p} \frac{d}{d\mathbf{x}_{tj}} Q^{-1}(\mathbf{x}_{tj}, \mathbf{z}_t; \theta^*_{tj}(\phi)) q(\mathbf{z}_{C+1:T}|\mathbf{x}_{1:C};\phi) d\mathbf{z}_{C+1:T} \\
=& \int \prod_{t=C+1}^{T} \prod_{j=1}^{p} q_{tj}(\mathbf{x}_{tj}|\mathbf{z}_t;\phi) q(\mathbf{z}_{C+1:T}|\mathbf{x}_{1:C};\phi) d\mathbf{z}_{C+1:T} \\
=& \int q(\mathbf{x}_{C+1:T}|\mathbf{z}_{C+1:T};\phi) q(\mathbf{z}_{C+1:T}|\mathbf{x}_{1:C};\phi) d\mathbf{z}_{C+1:T} \\
=& \int p(\mathbf{x}_{C+1:T}|\mathbf{x}_{1:C}) q(\mathbf{z}_{C+1:T}|\mathbf{x}_{1:T};\phi) d\mathbf{z}_{C+1:T} \\
=& p(\mathbf{x}_{C+1:T}|\mathbf{x}_{1:C}) \int q(\mathbf{z}_{C+1:T}|\mathbf{x}_{1:T};\phi) d\mathbf{z}_{C+1:T} \\
=& p(\mathbf{x}_{C+1:T}|\mathbf{x}_{1:C}).
\end{aligned}$$

The proof is complete. □

B. Proof of Lemma 2

Lemma. *For an arbitrary ϕ, let $\theta^*(\phi) \in \arg\min_\theta \mathbb{E}_p \mathbb{E}_q \sum_{t=C+1}^{T} \sum_{j=1}^{p} \int_0^1 \rho_\alpha\big(\mathbf{x}_{tj} - Q(\alpha, \mathbf{z}_t; \theta_{tj})\big) d\alpha$, where $\theta^*(\phi) = (\theta^*_{C+1}(\phi), \cdots, \theta^*_T(\phi))$, $\theta^*_t(\phi) = (\theta^*_{t1}(\phi), \cdots, \theta^*_{tp}(\phi))$, $t = C+1, \cdots, T$, and p and q denote $p(\mathbf{x}_{C+1:T}|\mathbf{x}_{1:C})$ and $q(\mathbf{z}_{C+1:T}|\mathbf{x}_{1:T};\phi)$, respectively. Define $h : \mathbb{R}^{(T-C)p} \mapsto \mathbb{R}_+$ (vectorized) as*

$$h(\mathbf{x}_{C+1:T}|\mathbf{x}_{1:C};\phi,\eta) := \int \frac{p(\mathbf{z}_{C+1:T}|\mathbf{x}_{1:C};\eta) q(\mathbf{z}_{C+1:T}|\mathbf{x}_{1:T};\phi)}{q(\mathbf{z}_{C+1:T}|\mathbf{x}_{1:C};\phi)} d\mathbf{z}_{C+1:T}$$

with assuming continuity, and define the estimated pdf $\hat{p}(\mathbf{x}_{C+1:T}|\mathbf{x}_{1:C};\theta^*(\phi),\eta)$ as

$$\hat{p}(\mathbf{x}_{C+1:T}|\mathbf{x}_{1:C};\theta^*(\phi),\eta) := \int \prod_{t=1}^{T} \prod_{j=1}^{p} \frac{d}{d\mathbf{x}_{tj}} Q^{-1}(\mathbf{x}_{tj}, \mathbf{z}_t; \theta^*_{tj}(\phi)) p(\mathbf{z}_{C+1:T}|\mathbf{x}_{1:C};\eta) d\mathbf{z}_{C+1:T}.$$

Suppose that $q(\mathbf{x}_{C+1:T}|\mathbf{z}_{C+1:T};\phi)p(\mathbf{z}_{C+1:T}|\mathbf{x}_{1:C};\eta)$ is the non-negative measurable function. Then, there exists $\mathbf{a}\in\mathbb{R}^{(T-C)\times p}$ such that $h(\mathbf{a}|\mathbf{x}_{1:C};\phi,\eta)=1$. Furthermore, if $h(\cdot|\mathbf{x}_{1:C};\phi,\eta)$ is twice continuously differentiable in a closed ball $B=\{\mathbf{u}\in\mathbb{R}^{(T-C)\times p}:\|\mathbf{a}-\mathbf{u}\|\leq r\}$ for some $r>0$ and the hessian matrix of $h(\mathbf{u}|\mathbf{x}_{1:C};\phi,\eta)\log h(\mathbf{u}|\mathbf{x}_{1:C};\phi,\eta)$, $H(\mathbf{u}|\mathbf{x}_{1:C};\phi,\eta)$, satisfies $\max_{\mathbf{u}\in B}\|H(\mathbf{u}|\mathbf{x}_{1:C};\phi,\eta)\|_{\max}\leq M(\mathbf{x}_{1:C})_{\phi,\eta}<\infty$, then

$$
\begin{aligned}
&\mathcal{KL}\Big(\hat{p}(\mathbf{x}_{C+1:T}|\mathbf{x}_{1:C};\theta^*(\phi),\eta)\|p(\mathbf{x}_{C+1:T}|\mathbf{x}_{1:C})\Big)\\
=&\mathcal{KL}\Big(p(\mathbf{x}_{C+1:T}|\mathbf{x}_{1:C})h(\mathbf{x}_{C+1:T}|\mathbf{x}_{1:C};\phi,\eta)\|p(\mathbf{x}_{C+1:T}|\mathbf{x}_{1:C})\Big)\\
\leq&\mathbb{E}_p\left[\nabla h(\mathbf{a}|\mathbf{x}_{1:C};\phi,\eta)^\top(vec(\mathbf{x}_{C+1:T})-vec(\mathbf{a}))+\frac{M(\mathbf{x}_{1:C})_{\phi,\eta}}{2}\|vec(\mathbf{x}_{C+1:T})-vec(\mathbf{a})\|^2\right],
\end{aligned}
$$

where $vec(\cdot)$ and p denote a flattening map and $p(\mathbf{x}_{C+1:T}|\mathbf{x}_{1:C})$ respectively.

Proof. Since $q(\mathbf{x}_{C+1:T}|\mathbf{z}_{C+1:T};\phi)p(\mathbf{z}_{C+1:T}|\mathbf{x}_{1:C};\eta)$ is the non-negative measurable function,

$$
\begin{aligned}
&\int p(\mathbf{x}_{C+1:T}|\mathbf{x}_{1:C})h(\mathbf{x}_{C+1:T}|\mathbf{x}_{1:C};\phi)d\mathbf{x}_{C+1:T}\\
=&\int p(\mathbf{x}_{C+1:T}|\mathbf{x}_{1:C})\int\frac{p(\mathbf{z}_{C+1:T}|\mathbf{x}_{1:C};\eta)q(\mathbf{z}_{C+1:T}|\mathbf{x}_{1:T};\phi)}{q(\mathbf{z}_{C+1:T}|\mathbf{x}_{1:C};\phi)}d\mathbf{z}_{C+1:T}d\mathbf{x}_{C+1:T}\\
=&\int\int q(\mathbf{x}_{C+1:T}|\mathbf{z}_{C+1:T};\phi)p(\mathbf{z}_{C+1:T}|\mathbf{x}_{1:C};\eta)d\mathbf{z}_{C+1:T}d\mathbf{x}_{C+1:T}\\
=&\int\int q(\mathbf{x}_{C+1:T}|\mathbf{z}_{C+1:T};\phi)p(\mathbf{z}_{C+1:T}|\mathbf{x}_{1:C};\eta)d\mathbf{x}_{C+1:T}d\mathbf{z}_{C+1:T}\\
=&\int p(\mathbf{z}_{C+1:T}|\mathbf{x}_{1:C};\eta)d\mathbf{z}_{C+1:T}\\
=&1,
\end{aligned}
$$

by Tonelli's theorem. Subtracting $\int p(\mathbf{x}_{C+1:T}|\mathbf{x}_{1:C})d\mathbf{x}_{C+1:T}=1$ from $\int p(\mathbf{x}_{C+1:T}|\mathbf{x}_{1:C})h(\mathbf{x}_{C+1:T}|\mathbf{x}_{1:C};\phi,\eta)d\mathbf{x}_{C+1:T}=1$ yields $\int p(\mathbf{x}_{C+1:T}|\mathbf{x}_{1:C})(h(\mathbf{x}_{C+1:T}|\mathbf{x}_{1:C};\phi,\eta)-1)d\mathbf{x}_{C+1:T}=0$. Since $p(\mathbf{x}_{C+1:T}|\mathbf{x}_{1:C})>0$ for all $\mathbf{x}\in\mathbb{R}^{(T-C)\times p}$,

- if $h(\mathbf{x}_{C+1:T}|\mathbf{x}_{1:C};\phi,\eta)>1$ for all $\mathbf{x}\in\mathbb{R}^{(T-C)\times p}$, then $\int p(\mathbf{x}_{C+1:T}|\mathbf{x}_{1:C})(h(\mathbf{x}_{C+1:T}|\mathbf{x}_{1:C};\phi,\eta)-1)d\mathbf{x}_{C+1:T}>0$, and
- if $h(\mathbf{x}_{C+1:T}|\mathbf{x}_{1:C};\phi,\eta)<1$ for all $\mathbf{x}\in\mathbb{R}^{(T-C)\times p}$, then $\int p(\mathbf{x}_{C+1:T}|\mathbf{x}_{1:C})(h(\mathbf{x}_{C+1:T}|\mathbf{x}_{1:C};\phi,\eta)-1)d\mathbf{x}_{C+1:T}<0$.

Therefore, $\int p(\mathbf{x}_{C+1:T}|\mathbf{x}_{1:C})(h(\mathbf{x}_{C+1:T}|\mathbf{x}_{1:C};\phi,\eta)-1)d\mathbf{x}_{C+1:T}=0$ and the continuity of $h(\cdot|\mathbf{x}_{1:C};\phi,\eta)$ implies that there exists $\mathbf{a}\in\mathbb{R}^p$ such that $h(\mathbf{a}|\mathbf{x}_{1:C};\phi,\eta)=1$.

$$
\begin{aligned}
\min_{\theta}\quad&\mathbb{E}_{p(\mathbf{x}_{C+1:T}|\mathbf{x}_{1:C})}\mathbb{E}_{q(\mathbf{z}_{C+1:T}|\mathbf{x}_{1:T};\phi)}\sum_{t=C+1}^{T}\sum_{j=1}^{p}\int_0^1\rho_\alpha\big(\mathbf{x}_{tj}-Q(\alpha,\mathbf{z}_t;\theta_{tj})\big)d\alpha\\
&=\mathbb{E}_{q(\mathbf{x}_{C+1:T}|\mathbf{z}_{C+1:T};\phi)q(\mathbf{z}_{C+1:T}|\mathbf{x}_{1:C};\phi)}\sum_{t=C+1}^{T}\sum_{j=1}^{p}\int_0^1\rho_\alpha\big(\mathbf{x}_{tj}-Q(\alpha,\mathbf{z}_t;\theta_{tj})\big)d\alpha\\
&=\mathbb{E}_{q(\mathbf{z}_{C+1:T}|\mathbf{x}_{1:C};\phi)}\sum_{t=C+1}^{T}\sum_{j=1}^{p}\mathbb{E}_{q(\mathbf{x}_{tj}|\mathbf{z}_t;\phi)}\int_0^1\rho_\alpha\big(\mathbf{x}_{tj}-Q(\alpha,\mathbf{z}_t;\theta_{tj})\big)d\alpha
\end{aligned}
$$

Therefore, for all $t=C+1,\cdots,T$, and $j=1,\cdots,p$,

$$
\theta^*_{tj}(\phi)=\arg\min_{\theta_{tj}}\mathbb{E}_{q(\mathbf{x}_{tj}|\mathbf{z}_t;\phi)}\int_0^1\rho_\alpha\big(\mathbf{x}_{tj}-Q(\alpha,\mathbf{z}_t;\theta_{tj})\big)d\alpha
$$

and it is proper relative to $F_{tj}(\mathbf{x}_{tj}|\mathbf{z}_t;\phi)$ (CRPS loss). It implies that $Q^{-1}(\mathbf{x}_{tj},\mathbf{z}_t;\theta^*_{tj}(\phi))=F_{tj}(\mathbf{x}_{tj}|\mathbf{z}_t;\phi)$, and

$\frac{d}{d\mathbf{x}_{tj}}Q^{-1}(\mathbf{x}_{tj}|\mathbf{z}_t;\theta^*_{tj}(\phi)) = q_{tj}(\mathbf{x}_{tj}|\mathbf{z}_t;\phi)$, by assumption. And

$$\begin{aligned}\hat{p}(\mathbf{x}_{C+1:T}|\mathbf{x}_{1:C};\theta^*(\phi))=&\int\prod_{t=1}^{T}\prod_{j=1}^{p}\frac{d}{d\mathbf{x}_{tj}}Q^{-1}(\mathbf{x}_{tj},\mathbf{z}_t;\theta^*_{tj}(\phi))p(\mathbf{z}_{C+1:T}|\mathbf{x}_{1:C};\eta)d\mathbf{z}_{C+1:T}\\ =&\int\prod_{t=1}^{T}\prod_{j=1}^{p}q_{tj}(\mathbf{x}_{tj}|\mathbf{z}_t;\phi)p(\mathbf{z}_{C+1:T}|\mathbf{x}_{1:C};\eta)d\mathbf{z}_{C+1:T}\\ =&\int q(\mathbf{x}_{C+1:T}|\mathbf{z}_{C+1:T};\phi)p(\mathbf{z}_{C+1:T}|\mathbf{x}_{1:C};\eta)d\mathbf{z}_{C+1:T}\\ =&\int\frac{p(\mathbf{x}_{C+1:T}|\mathbf{x}_{1:C})q(\mathbf{z}_{C+1:T}|\mathbf{x}_{1:T};\phi)}{q(\mathbf{z}_{C+1:T}|\mathbf{x}_{1:C};\phi)}p(\mathbf{z}_{C+1:T}|\mathbf{x}_{1:C};\eta)d\mathbf{z}_{C+1:T}\\ =&p(\mathbf{x}_{C+1:T}|\mathbf{x}_{1:C})\int\frac{p(\mathbf{z}_{C+1:T}|\mathbf{x}_{1:C};\eta)q(\mathbf{z}_{C+1:T}|\mathbf{x}_{1:T};\phi)}{q(\mathbf{z}_{C+1:T}|\mathbf{x}_{1:C};\phi)}d\mathbf{z}_{C+1:T}\\ =&p(\mathbf{x}_{C+1:T}|\mathbf{x}_{1:C})h(\mathbf{x}_{C+1:T}|\mathbf{x}_{1:C};\phi,\eta).\end{aligned}$$

Therefore,

$$\begin{aligned}&\mathcal{KL}\Big(\hat{p}(\mathbf{x}_{C+1:T}|\mathbf{x}_{1:C};\theta^*(\phi))\|p(\mathbf{x}_{C+1:T}|\mathbf{x}_{1:C})\Big)\\ =&\mathcal{KL}\Big(p(\mathbf{x}_{C+1:T}|\mathbf{x}_{1:C})h(\mathbf{x}_{C+1:T}|\mathbf{x}_{1:C};\phi,\eta)\|p(\mathbf{x}_{C+1:T}|\mathbf{x}_{1:C})\Big)\\ =&\int p(\mathbf{x}_{C+1:T}|\mathbf{x}_{1:C})h(\mathbf{x}_{C+1:T}|\mathbf{x}_{1:C};\phi,\eta)\log\frac{p(\mathbf{x}_{C+1:T}|\mathbf{x}_{1:C})h(\mathbf{x}_{C+1:T}|\mathbf{x}_{1:C};\phi,\eta)}{p(\mathbf{x}_{C+1:T}|\mathbf{x}_{1:C})}d\mathbf{x}_{C+1:T}\\ =&\mathbb{E}_{p(\mathbf{x}_{C+1:T}|\mathbf{x}_{1:C})}\left[h(\mathbf{x}_{C+1:T}|\mathbf{x}_{1:C};\phi,\eta)\log h(\mathbf{x}_{C+1:T}|\mathbf{x}_{1:C};\phi,\eta)\right]\\ \leq&\mathbb{E}_p\left[\nabla h(\mathbf{a}|\mathbf{x}_{1:C};\phi,\eta)^\top(vec(\mathbf{x}_{C+1:T})-vec(\mathbf{a}))+\frac{M(\mathbf{x}_{1:C})_{\phi,\eta}}{2}\|vec(\mathbf{x}_{C+1:T})-vec(\mathbf{a})\|^2\right],\end{aligned}$$

by Taylor's theorem and the assumptions. The proof is complete. □

C. *Latent Space Architectures*

For the model architectures of the latent space, we follow the construction of [16]. The parameters of prior distribution at time $t = C+1,\cdots,T$ can be described by following pseudocode (see Figure 2 for overall architectures):

$$\begin{aligned}\bar{\mathbf{w}}_t=&\mathtt{LN}(\mathbf{w}_{t-1}+\mathtt{Att}^{(1)}_t(\mathbf{w}_{t-1},\mathbf{w}_{1:t-1},\mathbf{w}_{1:t-1};\eta_t);\eta_t)\\ \hat{\mathbf{w}}_t=&\mathtt{LN}(\bar{\mathbf{w}}_t+\mathtt{Att}^{(2)}_t(\bar{\mathbf{w}}_t,\mathbf{h}_{1:C},\mathbf{h}_{1:C};\eta_t);\eta_t)\\ \mu(\mathbf{z}_{1:t-1},\mathbf{x}_{1:C};\eta_t)=&\mathtt{MLP}^{(1)}_t(\hat{\mathbf{w}}_t;\eta_t)\\ \mathbf{z}_t=&\mu(\mathbf{z}_{1:t-1},\mathbf{x}_{1:C};\eta_t)+s\cdot\boldsymbol{\epsilon}\\ \mathbf{w}_t=&\mathtt{LN}(\hat{\mathbf{w}}_t+\mathtt{MLP}^{(2)}_t(\mathbf{z}_t;\eta_t)+\mathtt{Pos}(t);\eta_t);\eta_t),\end{aligned}$$

where

$$\begin{aligned}\mathbf{h}_t=&\mathtt{LN}(\mathtt{MLP}^{(0)}_t(\mathbf{x}_t;\eta_t)+\mathtt{Pos}(t);\eta_t)\in\mathbb{R}^m\\ \mathbf{h}_{1:C}=&[\mathbf{h}_1,\cdots,\mathbf{h}_C]^\top\in\mathbb{R}^{C\times m}\\ \mathtt{Att}^{(1)}_t\colon&\ \mathbb{R}^m\times\mathbb{R}^{(t-1)\times m}\times\mathbb{R}^{(t-1)\times m}\mapsto\mathbb{R}^m\\ \mathtt{Att}^{(2)}_t\colon&\ \mathbb{R}^m\times\mathbb{R}^{C\times m}\times\mathbb{R}^{C\times m}\mapsto\mathbb{R}^m\\ \mathtt{MLP}^{(1)}_t\colon&\ \mathbb{R}^m\mapsto\mathbb{R}^d\\ \mathtt{MLP}^{(2)}_t\colon&\ \mathbb{R}^m\mapsto\mathbb{R}^m\\ \boldsymbol{\epsilon}\sim&N(0,I),\end{aligned}$$

`LN` and `Att` represent LayerNormalization and Attention layer, respectively, and a trainable vector $\mathbf{w}_0$ is initialized with $N(0, 0.01)$.

Since the parameters of posterior distribution at time t can also be summarized similarly with the prior distribution, we describe the only different parts due to the limitation of the space:

$$\begin{aligned}\mathbf{k}_t=&\mathtt{Att}^{(0)}_t(\mathbf{h}_t,\mathbf{h}_{1:T},\mathbf{h}_{1:T};\phi_t)\\ \big(\mu(\mathbf{z}_{1:t-1},\mathbf{x}_{1:T};\phi_t),\sigma^2(\mathbf{z}_{1:t-1},\mathbf{x}_{1:T};\phi_t)\big)=&\mathtt{MLP}^{(1)}_t(\hat{\mathbf{w}}_t,\mathbf{k}_t;\phi_t)\\ \mathbf{z}_t=&\mu(\mathbf{z}_{1:t-1},\mathbf{x}_{1:T};\phi_t)+\sigma^2(\mathbf{z}_{1:t-1},\mathbf{x}_{1:T};\phi_t)\cdot\boldsymbol{\epsilon},\end{aligned}$$

where

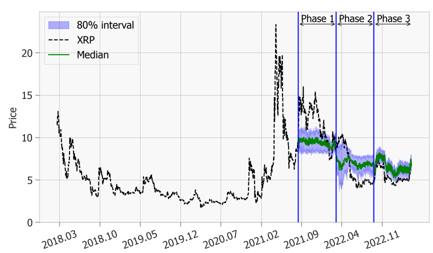
$$\mathbf{h}_{1:T} = [\mathbf{h}_1, \cdots, \mathbf{h}_T]^\top \in \mathbb{R}^{T \times m}$$
$$\mathtt{Att}_t^{(0)}: \mathbb{R}^m \times \mathbb{R}^{T \times m} \times \mathbb{R}^{T \times m} \mapsto \mathbb{R}^m$$
$$\mathtt{MLP}_t^{(1)}: \mathbb{R}^m \times \mathbb{R}^m \mapsto \mathbb{R}^{2d}.$$

D. Experimental Details

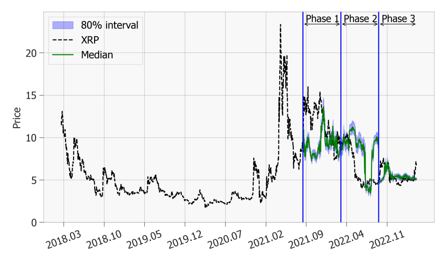
TABLE IV: Hyperparameter settings for $T - C = 1$. The hyperparameter in the parenthesis is used when $T - C = 5$.

Model	epochs	learning rate	batch size	d	m	# head	# layer	β	s^2	K
SQF-RNN	1000	0.001	100	-	30	-	3	-	-	-
DeepAR	1000	0.001	100	-	20	-	3	-	-	-
GP-Copula	1000	0.001	128	-	30	-	2	-	-	-
TLAE	1000	0.0001	256	32	-	-	4	0.001	-	-
ProTran	200	0.005	256	16	8	1	2	0.1	-	-
MQRNN	1000	0.001	100	-	20	-	3	-	-	-
TFT	1500	0.001	100	-	20	1	-	-	-	-
DMFVAE(LSQF)	200	0.0025	256	16	8	1	1	0.1	0.1	-
DMFVAE(ExpLog)	500	0.0005	256	16	8	1	1	0.5	0.1	20
DMFVAE(GLD, finite)	200	0.0025	256	16	8	1	1	0.1(0.1)	0.1(0.5)	20
DMFVAE(GLD, inf)	200	0.0025	256	16	8	1	1	0.1(0.1)	0.1(0.1)	20

E. XRP forecasting results of the GP-Copula and TFT



(a) GP-Copula



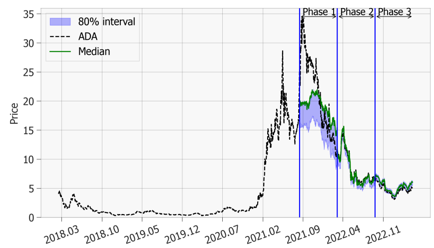
(b) TFT

Fig. 6: Visualization Results with the XRP. $T - C = 5$.

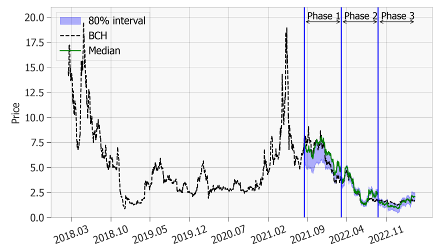
F. Full Results

TABLE V: $T - C = 1$. Normalized CRPS, Normalized Quantile loss. DMFVAE(GLD, finite). Lower is better.

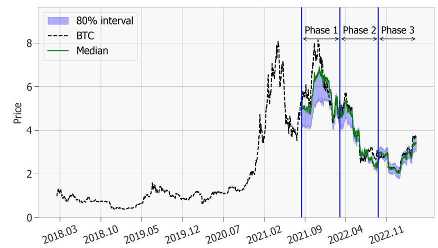
	1st Phase				2nd Phase				3rd Phase			
Target	CRPS	QL(0.1)	QL(0.5)	QL(0.9)	CRPS	QL(0.1)	QL(0.5)	QL(0.9)	CRPS	QL(0.1)	QL(0.5)	QL(0.9)
BTC	0.058	0.015	0.035	0.051	0.038	0.011	0.023	0.024	0.025	0.009	0.018	0.006
ETH	0.173	0.037	0.091	0.151	0.102	0.025	0.051	0.080	0.080	0.019	0.052	0.043
XRP	0.085	0.023	0.051	0.069	0.018	0.007	0.010	0.006	0.019	0.006	0.013	0.006
ADA	0.112	0.021	0.069	0.096	0.022	0.008	0.013	0.013	0.055	0.018	0.043	0.013
ETC	0.050	0.027	0.024	0.031	0.228	0.036	0.117	0.201	0.066	0.025	0.051	0.007
BCH	0.058	0.018	0.041	0.026	0.048	0.027	0.033	0.011	0.086	0.028	0.061	0.020
XLM	0.108	0.027	0.082	0.027	0.077	0.015	0.051	0.048	0.053	0.011	0.030	0.026
KOSPI	0.058	0.012	0.048	0.010	0.072	0.021	0.053	0.023	0.035	0.017	0.019	0.013
DJI	0.020	0.004	0.017	0.004	0.027	0.010	0.021	0.006	0.056	0.013	0.034	0.038
VIX	0.077	0.015	0.053	0.033	0.280	0.039	0.232	0.082	0.135	0.022	0.110	0.035



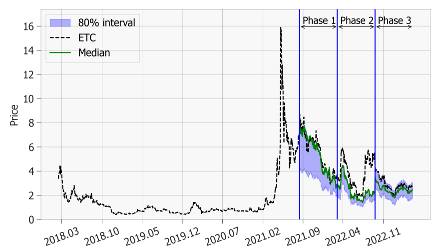
(a) ADA



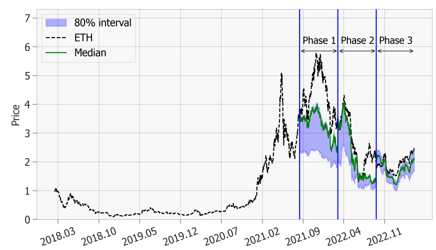
(b) BCH



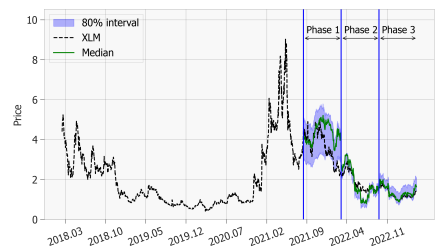
(c) BTC



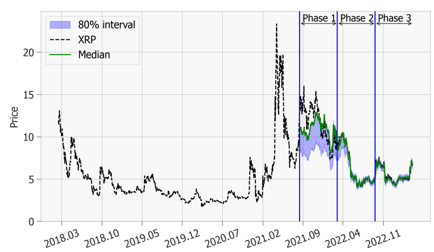
(d) ETC



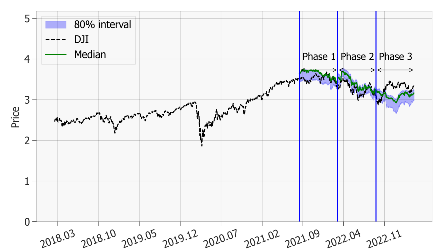
(e) ETH



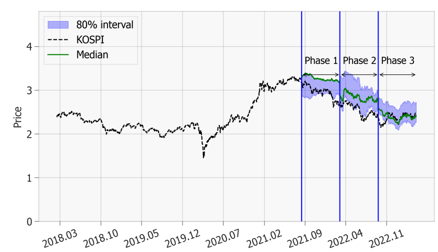
(f) XLM



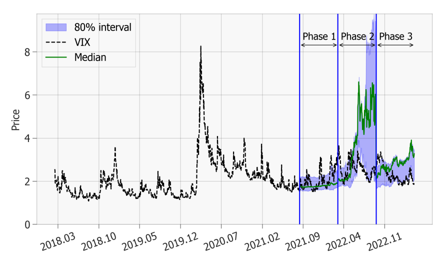
(g) XRP



(h) DJI



(i) KOSPI

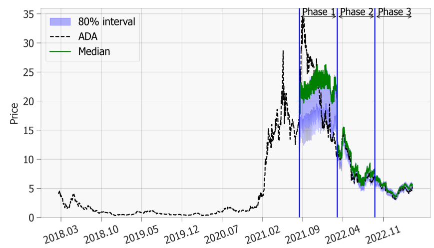


(j) VIX

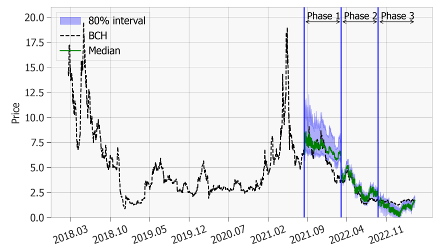
Fig. 7: $T - C = 1$. DMFVAE(GLD, finite).

TABLE VI: $T - C = 5$. CRPS, Quantile loss. DMFVAE(GLD, finite). Lower is better.

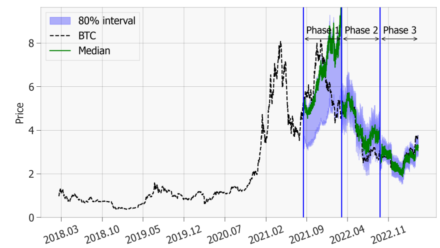
	1st Phase				2nd Phase				3rd Phase			
Target	CRPS	QL(0.1)	QL(0.5)	QL(0.9)	CRPS	QL(0.1)	QL(0.5)	QL(0.9)	CRPS	QL(0.1)	QL(0.5)	QL(0.9)
BTC	0.132	0.027	0.098	0.055	0.070	0.045	0.043	0.019	0.043	0.014	0.030	0.015
ETH	0.108	0.036	0.060	0.065	0.042	0.016	0.025	0.023	0.100	0.025	0.067	0.070
XRP	0.089	0.026	0.060	0.053	0.041	0.013	0.028	0.014	0.054	0.017	0.038	0.023
ADA	0.133	0.035	0.091	0.076	0.039	0.012	0.033	0.010	0.036	0.009	0.031	0.010
ETC	0.125	0.027	0.117	0.040	0.186	0.036	0.110	0.129	0.060	0.026	0.043	0.015
BCH	0.098	0.054	0.056	0.030	0.089	0.052	0.059	0.021	0.185	0.044	0.140	0.079
XLM	0.219	0.101	0.157	0.039	0.111	0.046	0.075	0.030	0.106	0.028	0.078	0.032
KOSPI	0.037	0.015	0.030	0.007	0.058	0.007	0.041	0.028	0.027	0.010	0.016	0.010
DJI	0.022	0.010	0.017	0.005	0.026	0.008	0.019	0.008	0.054	0.014	0.034	0.033
VIX	0.098	0.018	0.059	0.063	0.100	0.021	0.071	0.047	0.102	0.017	0.083	0.031



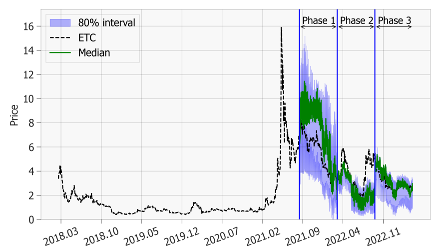
(a) ADA



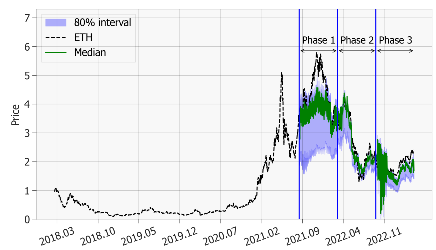
(b) BCH



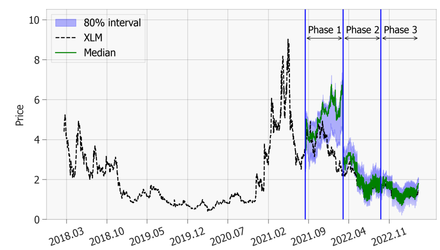
(c) BTC



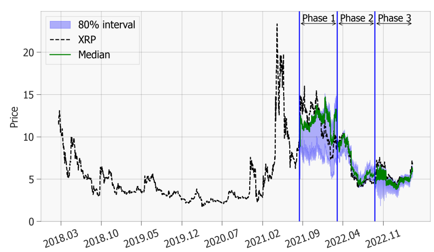
(d) ETC



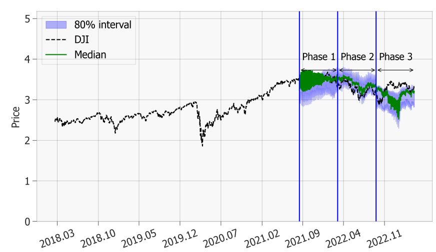
(e) ETH



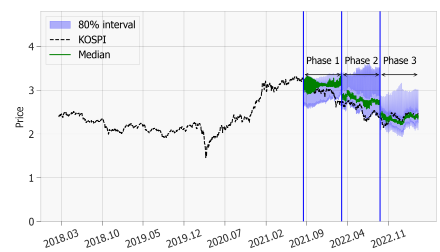
(f) XLM



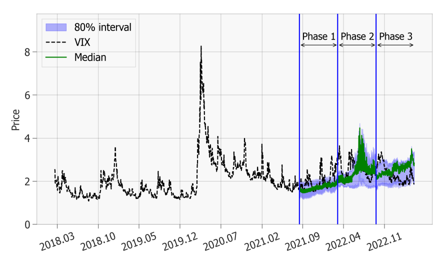
(g) XRP



(h) DJI



(i) KOSPI

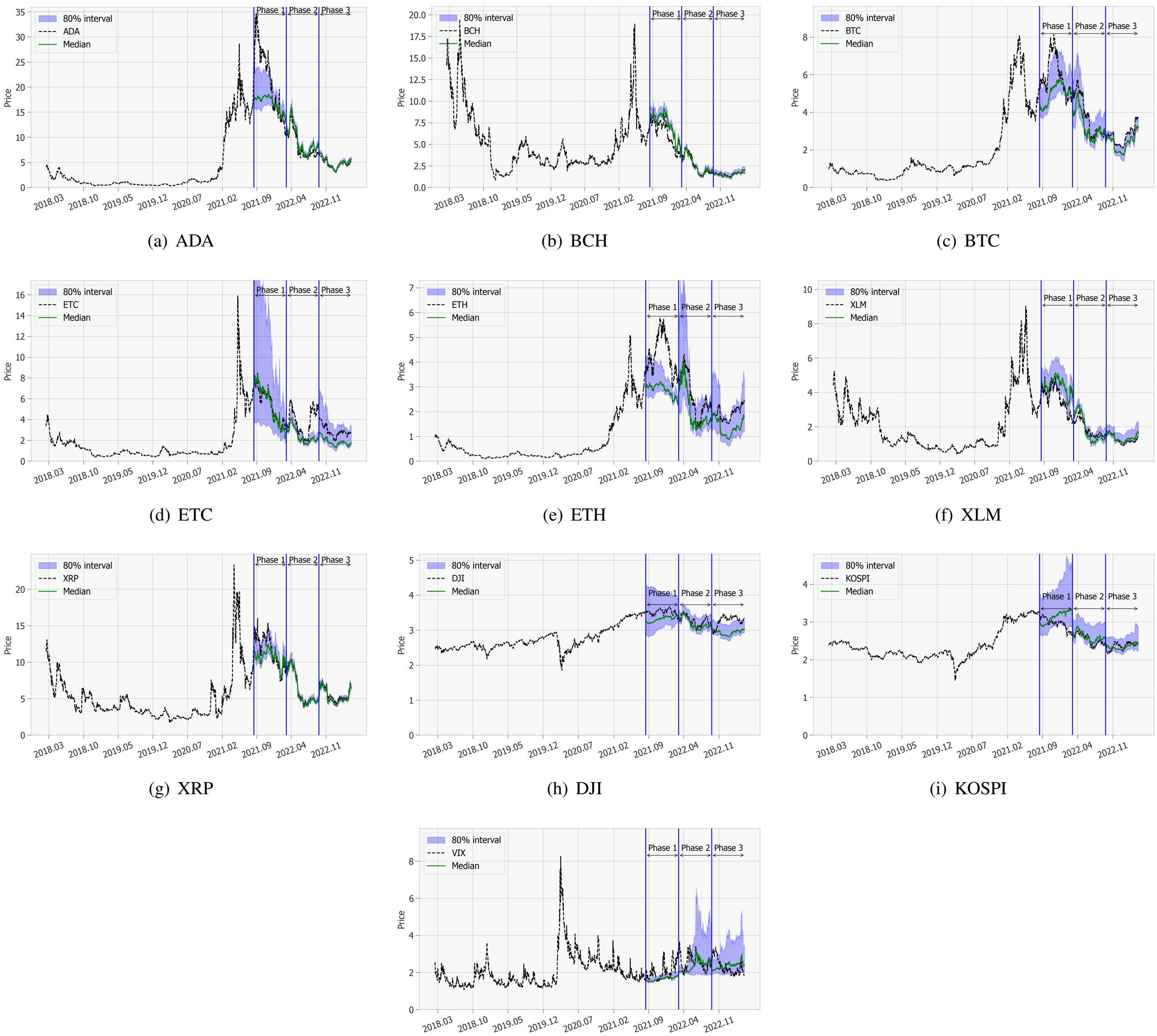


(j) VIX

Fig. 8: $T - C = 5$. DMFVAE(GLD, finite).

TABLE VII: $T - C = 1$. DICR. DMFVAE(GLD, finite). Lower is better.

	1st Phase		2nd Phase		3rd Phase	
Model	DICR	MI	DICR	MI	DICR	MI
BTC	0.430	0.110	0.515	0.080	0.060	0.133
ETH	0.800	0.200	0.605	0.170	0.545	0.172
XRP	0.360	0.182	0.020	0.094	0.070	0.089
ADA	0.325	0.141	0.430	0.071	0.405	0.121
ETC	0.475	0.253	0.795	0.132	0.060	0.298
BCH	0.270	0.186	0.150	0.131	0.105	0.365
XLM	0.005	0.402	0.330	0.136	0.025	0.294
KOSPI	0.005	0.124	0.555	0.210	0.045	0.171
DJI	0.085	0.059	0.390	0.060	0.650	0.083
VIX	0.090	0.205	0.225	0.772	0.265	0.313



(a) ADA (b) BCH (c) BTC

(d) ETC (e) ETH (f) XLM

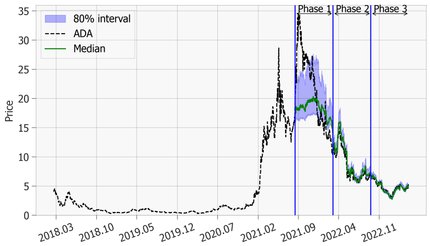
(g) XRP (h) DJI (i) KOSPI

(j) VIX

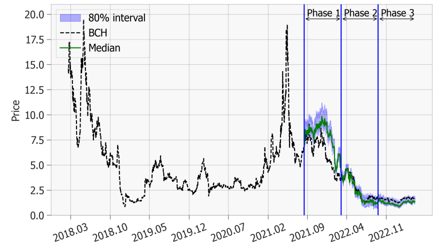
Fig. 9: $T - C = 1$. DMFVAE(GLD, inf).

TABLE VIII: $T - C = 5$. DICR. DMFVAE(GLD, finite). Lower is better.

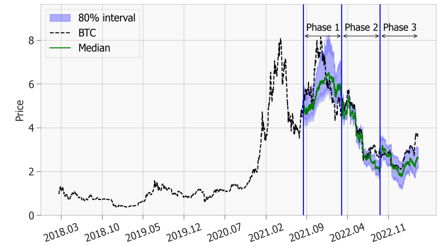
	1^{st} Phase		2^{nd} Phase		3^{rd} Phase	
Model	DICR	MI	DICR	MI	DICR	MI
BTC	0.437	0.318	0.313	0.192	0.075	0.180
ETH	0.432	0.311	0.382	0.143	0.520	0.216
XRP	0.296	0.308	0.306	0.151	0.254	0.166
ADA	0.583	0.184	0.077	0.119	0.020	0.150
ETC	0.124	0.636	0.529	0.232	0.080	0.351
BCH	0.249	0.318	0.472	0.169	0.380	0.418
XLM	0.251	0.367	0.344	0.310	0.331	0.356
KOSPI	0.024	0.138	0.206	0.289	0.012	0.157
DJI	0.010	0.118	0.143	0.090	0.500	0.119
VIX	0.293	0.151	0.056	0.314	0.118	0.305



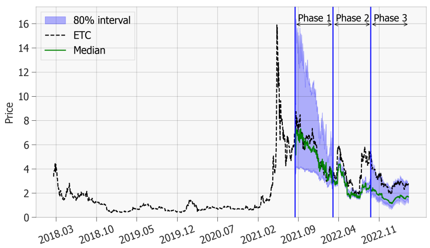
(a) ADA



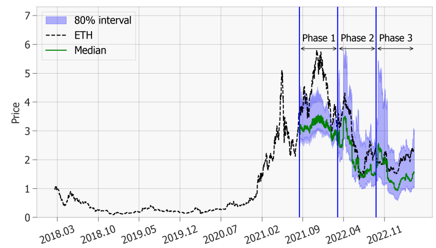
(b) BCH



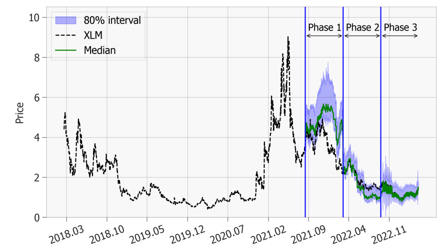
(c) BTC



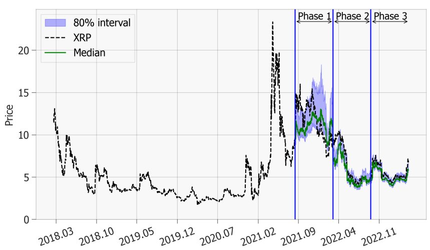
(d) ETC



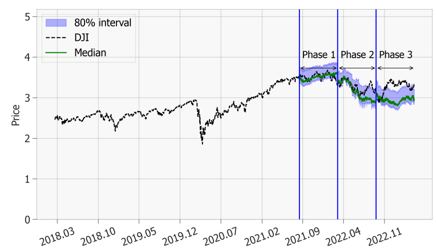
(e) ETH



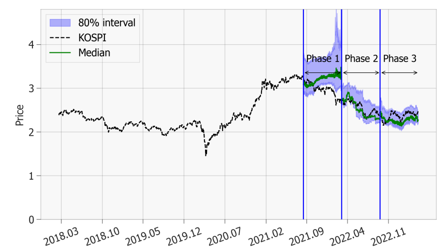
(f) XLM



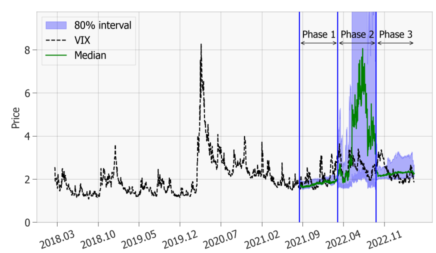
(g) XRP



(h) DJI



(i) KOSPI



(j) VIX

Fig. 10: $T - C = 5$. DMFVAE(GLD, inf).

TABLE IX: $T - C = 1$. CRPS, Quantile loss. DMFVAE(GLD, infinite). Lower is better.

	1st Phase				2nd Phase				3rd Phase			
Target	CRPS	QL(0.1)	QL(0.5)	QL(0.9)	CRPS	QL(0.1)	QL(0.5)	QL(0.9)	CRPS	QL(0.1)	QL(0.5)	QL(0.9)
BTC	0.093	0.018	0.067	0.033	0.059	0.016	0.046	0.016	0.060	0.015	0.050	0.012
ETH	0.201	0.030	0.126	0.128	0.092	0.025	0.063	0.028	0.145	0.038	0.114	0.021
XRP	0.061	0.014	0.045	0.016	0.016	0.005	0.011	0.005	0.023	0.007	0.020	0.006
ADA	0.116	0.023	0.077	0.058	0.051	0.040	0.030	0.011	0.031	0.017	0.020	0.009
ETC	0.098	0.028	0.026	0.073	0.186	0.026	0.114	0.126	0.148	0.031	0.119	0.026
BCH	0.104	0.082	0.059	0.019	0.031	0.011	0.021	0.009	0.052	0.019	0.031	0.020
XLM	0.118	0.051	0.077	0.034	0.073	0.040	0.051	0.015	0.055	0.009	0.033	0.031
KOSPI	0.063	0.013	0.044	0.032	0.027	0.004	0.016	0.016	0.025	0.008	0.016	0.011
DJI	0.037	0.013	0.028	0.016	0.017	0.005	0.013	0.006	0.066	0.013	0.050	0.026
VIX	0.106	0.015	0.064	0.074	0.097	0.019	0.052	0.066	0.097	0.015	0.065	0.052

TABLE X: $T - C = 5$. CRPS, Quantile loss. DMFVAE(GLD, infinite). Lower is better.

	1st Phase				2nd Phase				3rd Phase			
Target	CRPS	QL(0.1)	QL(0.5)	QL(0.9)	CRPS	QL(0.1)	QL(0.5)	QL(0.9)	CRPS	QL(0.1)	QL(0.5)	QL(0.9)
BTC	0.083	0.024	0.058	0.030	0.049	0.013	0.034	0.020	0.075	0.021	0.053	0.025
ETH	0.159	0.027	0.105	0.088	0.090	0.025	0.071	0.018	0.142	0.037	0.107	0.040
XRP	0.086	0.025	0.061	0.028	0.058	0.012	0.043	0.024	0.038	0.010	0.032	0.009
ADA	0.119	0.030	0.078	0.053	0.046	0.029	0.029	0.012	0.029	0.010	0.021	0.008
ETC	0.074	0.023	0.032	0.048	0.169	0.023	0.103	0.124	0.143	0.029	0.111	0.010
BCH	0.162	0.124	0.092	0.029	0.041	0.014	0.029	0.011	0.097	0.027	0.090	0.013
XLM	0.203	0.127	0.118	0.047	0.069	0.017	0.052	0.018	0.047	0.015	0.028	0.022
KOSPI	0.070	0.031	0.043	0.026	0.024	0.007	0.016	0.011	0.032	0.009	0.026	0.006
DJI	0.016	0.006	0.010	0.007	0.027	0.006	0.021	0.008	0.067	0.013	0.049	0.026
VIX	0.086	0.018	0.052	0.055	0.300	0.021	0.230	0.200	0.084	0.020	0.054	0.045

TABLE XI: DICR. DMFVAE(GLD, infinite). Lower is better.

	$T - C = 1$						$T - C = 5$					
	1st Phase		2nd Phase		3rd Phase		1st Phase		2nd Phase		3rd Phase	
Model	DICR	MI	DICR	MI	DICR	MI	DICR	MI	DICR	MI	DICR	MI
BTC	0.310	0.219	0.085	0.289	0.340	0.151	0.439	0.190	0.305	0.117	0.045	0.271
ETH	0.620	0.164	0.060	0.464	0.175	0.589	0.594	0.180	0.176	0.421	0.172	0.765
XRP	0.255	0.161	0.095	0.081	0.090	0.099	0.373	0.240	0.105	0.136	0.021	0.122
ADA	0.500	0.153	0.660	0.065	0.240	0.094	0.498	0.199	0.465	0.097	0.106	0.106
ETC	0.200	1.008	0.465	0.134	0.180	0.561	0.189	0.711	0.601	0.102	0.004	0.324
BCH	0.760	0.104	0.115	0.107	0.105	0.299	0.728	0.153	0.000	0.176	0.010	0.329
XLM	0.420	0.302	0.135	0.208	0.020	0.349	0.558	0.349	0.050	0.276	0.121	0.344
KOSPI	0.020	0.368	0.090	0.195	0.105	0.164	0.466	0.251	0.086	0.159	0.107	0.138
DJI	0.200	0.289	0.200	0.110	0.460	0.117	0.071	0.096	0.027	0.106	0.473	0.117
VIX	0.440	0.081	0.040	0.483	0.115	0.471	0.483	0.086	0.193	2.205	0.257	0.260

TABLE XII: Standard deviations of metrics across 10 assets when $T - C = 1$.

	1st Phase				2nd Phase				3rd Phase			
Model	CRPS	QL(0.1)	QL(0.5)	QL(0.9)	CRPS	QL(0.1)	QL(0.5)	QL(0.9)	CRPS	QL(0.1)	QL(0.5)	QL(0.9)
SQF-RNN	0.150	0.133	0.081	0.030	0.095	0.079	0.048	0.066	0.068	0.059	0.036	0.029
DeepAR	0.105	0.081	0.068	0.038	0.069	0.065	0.036	0.041	0.061	0.032	0.042	0.032
GP-Copula	0.059	0.023	0.031	0.054	0.092	0.041	0.073	0.035	0.314	0.160	0.181	0.204
TLAE	0.203	0.017	0.100	0.193	0.249	0.206	0.141	0.061	0.206	0.083	0.146	0.046
ProTran	0.105	0.101	0.055	0.060	0.134	0.127	0.067	0.078	0.332	0.297	0.173	0.066
MQRNN	-	0.118	0.074	0.028	-	0.056	0.037	0.054	-	0.053	0.040	0.012
TFT	-	0.124	0.074	0.029	-	0.101	0.059	0.035	-	0.075	0.053	0.021
DMFVAE(LSQF)	0.045	0.007	0.030	0.032	0.050	0.012	0.030	0.039	0.031	0.011	0.022	0.016
DMFVAE(ExpLog)	0.025	0.010	0.043	0.038	0.067	0.014	0.058	0.071	0.041	0.009	0.055	0.050
DMFVAE(GLD, finite)	0.043	0.009	0.024	0.045	0.091	0.011	0.068	0.061	0.034	0.007	0.028	0.014
DMFVAE(GLD, inf)	0.044	0.022	0.029	0.035	0.052	0.014	0.031	0.038	0.046	0.010	0.038	0.013

TABLE XIII: Standard deviations of metrics across 10 assets when $T - C = 5$.

	1st Phase				2nd Phase				3rd Phase			
Model	CRPS	QL(0.1)	QL(0.5)	QL(0.9)	CRPS	QL(0.1)	QL(0.5)	QL(0.9)	CRPS	QL(0.1)	QL(0.5)	QL(0.9)
SQF-RNN	0.090	0.085	0.050	0.038	0.101	0.089	0.052	0.041	0.092	0.080	0.048	0.021
DeepAR	0.060	0.026	0.033	0.060	0.065	0.062	0.036	0.031	0.050	0.043	0.029	0.024
GP-Copula	0.240	0.030	0.134	0.190	0.083	0.037	0.057	0.039	0.287	0.151	0.167	0.174
TLAE	0.089	0.030	0.047	0.082	0.148	0.106	0.091	0.034	0.128	0.044	0.092	0.036
ProTran	0.076	0.068	0.040	0.064	0.136	0.130	0.069	0.072	0.366	0.338	0.186	0.078
MQRNN	-	0.135	0.091	0.031	-	0.052	0.033	0.029	-	0.092	0.064	0.017
TFT	-	0.037	0.032	0.061	-	0.112	0.061	0.031	-	0.111	0.070	0.016
DMFVAE(LSQF)	0.051	0.016	0.036	0.042	0.066	0.009	0.039	0.052	0.033	0.009	0.024	0.020
DMFVAE(ExpLog)	0.030	0.011	0.042	0.054	0.035	0.013	0.039	0.055	0.035	0.011	0.046	0.044
DMFVAE(GLD, finite)	0.054	0.026	0.042	0.024	0.048	0.017	0.028	0.035	0.048	0.010	0.037	0.024
DMFVAE(GLD, inf)	0.055	0.044	0.034	0.022	0.086	0.008	0.064	0.065	0.042	0.009	0.034	0.013